Facial Expression Analysis and The PAD Space

Varun Jain, James L. Crowley, Augustin Lux

FACIAL EXPRESSION ANALYSIS AND THE PAD SPACE

V. Jain[1, 2], J.L. Crowley[1, 3], A. Lux[1, 4]

[1]INRIA Grenoble Research Center, France
[2]varun.jain@inria.fr, [3]james.crowley@inria.fr, [4]augustin.lux@inria.fr

In this paper we present a technique for facial expression analysis and representing the underlying emotions in the PAD (Pleasure-Arousal-Dominance) space. We develop a purely appearance based approach using Multi-scale Gaussian derivatives and Support Vector Machines. The system can generalize well and is shown to outperform the baseline method.

Introduction

Facial expressions are a mirror to human emotions and are an important component of human to human interaction. Human computer interaction requires the same ability to read emotions from facial expressions.

Ekman introduced the concept of six basic emotions that are universally recognizable [1]. In [2] he presented the Facial Action Coding System (FACS): a taxonomy to describe facial expressions in terms of individual muscle movements.

FACS based approaches have been adopted in a variety of vision systems such as the Computer Expression Recognition Toolbox (CERT) [3]. Such systems are trained to estimate the Action Unit (AU) intensities which can then be used to assign one of the six basic emotion labels to that image or frame. The problem arises when the expression in the image is not associated with any of the six basic emotions.

An alternative to such a structured approach is to represent the underlying emotions in a multidimensional emotion space. One such method was presented by Dahmane and Meunier in [4]. The authors used Gabor wavelets and Support Vector Machines and represent the emotions that underlie the facial expressions using 4 dimensions (Activation, Expectation, Power and Valence).

In [5] the authors argue that three dimensions are enough to represent any emotion. In this paper we use the PAD emotional state model developed by Russell and Mehrabian and compare our results for Pleasure and Arousal with the results from the technique presented in [4].

Two common ways to describe image features are: appearance based methods and geometric feature based methods. The latter involves detection and tracking of facial keypoints such as the lip corners, nostrils and eyes which is done with the help of computationally expensive vision techniques and are not very robust.

The approach we present here does not involve identification of any landmarks on the face and just like the appearance based technique discussed in [4], the image filters are applied to the whole-face to obtain the feature vector.

PAD Emotional Space and datasets used

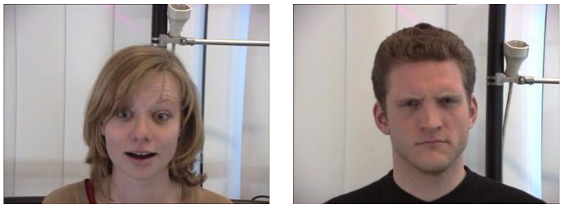
The PAD emotional state model is a psychological model developed by Russell and Mehrabian [5] that can be used to describe the emotional state of a person. Experiments support that 3 dimensions are sufficient to most human emotions.

The Pleasure-Displeasure Scale measures the pleasantness of an emotion while the Arousal-Nonarousal Scale measures the intensity of the emotion and the Dominance-Submissiveness Scale represents the controlling and dominant nature of the emotion.

The third axis of Dominance remains controversial and there is evidence to suggest that there is a high correlation between dominance and the other two axes [6].

Our approach was tested on the Cohn-Kanade [7] and FEED [8] datasets. The FEED dataset was collected at the Technical University of

Munich. The dataset was generated as a part of the European Union FG-NET project [9].
The FEED dataset does not contain posed emotions; the emotions were elicited by showing video clips to the participants. The database contains images from 18 individuals for 6 basic emotions along with the neutral face.



(a)Surprised (b) Angry

Fig. 1. Example Images from the FEED dataset

We map the basic emotions to the PAD space as shown in table 1 in accordance with the PAD values provided by Mehrabian. Instead of using numerical values we assign class labels (+P -P, +A -A) to perform binary classification.

Table 1. Labels for the 6 basic emotions

Emotion	P Label	A Label
Joy	+	+
Sadness	-	-
Surprise	+/-	+
Anger	-	+
Disgust	-	+
Fear	-	+

The Cohn-Kanade and FEED databases were re-annotated with these class labels. The Cohn-Kanade database was used for training and validation while the FEED database was used for testing.

Multi-scale Gaussian Derivatives

Gaussian derivatives can efficiently describe the neighborhood appearance of a pixel for pattern recognition tasks [10]. This is done by calculating different orders of Gaussian derivatives normalized in scale and orientation at every pixel.
The basic Gaussian function is defined as:

$$G(x,y;\sigma) = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Here σ is the scale factor or variance and defines the spatial support. This function measures the intensity of the neighborhood and does not contribute to the identification of the neighborhood and can be omitted.
First order derivatives provide information about the gradient (intensity and direction) whereas the second order derivatives provide the information about image features such as bars, blobs and corners. Higher order derivatives are only useful if the second order derivatives are strong otherwise they just contain image noise.
Obtaining scale invariant features is not a trivial task. Several methods have come up in the past addressing this problem. Lindeberg in [11] suggests that Gaussian derivatives be calculated across scales to get scale invariant features and then Lowe in [12] defines the intrinsic or characteristic scale as the value of the scale parameter at which the Laplacian provides a local maximum. The computational cost of directly searching the scale axis for this characteristic scale can be prohibitively expensive.
A cost-effective method for computing Multi-scale Gaussian derivatives has been discussed in detail in [13].
The next section is about Principal Component Analysis (PCA) and why we need it.

Principal Component Analysis

The region of the image containing the face is normalized to 64 X 64 pixels; this particular size is chosen after extensive experimentation where normalized images of 64 X 64 pixels gave the best accuracy. We calculate several orders of derivatives at 2 levels of scale for every pixel but it leads to an enormous feature vector therefore the image into cells of 4 X 4 pixels and the feature vector contains the mean and standard deviation of the descriptor values (Gaussian derivatives) for each cell of 4 X 4 pixels.

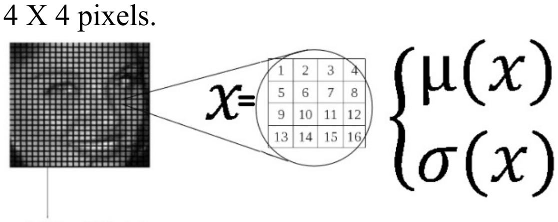


Fig. 2. The image divided into cells of 4 X 4 pixels.

Principal Component Analysis is used for dimensionality reduction which reduces the prediction time when the Support Vector Machines are used for classification.

PCA is done by eigenvalue decomposition of the data correlation matrix after normalizing the data for each dimension [14]. PCA provides us with scores and loadings. The scores are the transformed values corresponding to the data point and loadings are the coefficients the original variable should be multiplied with to get the score.

Support Vector Machines

Support Vector Machines (SVM) belong to a family of non-probabilistic linear classifiers [15]. The Radial Basis kernel provides the best accuracy for the particular application and is represented by the following equation:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (2)$$

We use a soft margin SVM, soft margin SVM's are used when the classes are not separable even after transforming the data to a higher dimension.

The Approach

Face detection is performed on the images in the dataset using the OpenCV face detector [16]. Following that a half-octave gaussian pyramid is constructed over a normalized imagette of the face. This is followed by dimensionality reduction by PCA and regression using Support Vector Machines.

Results

We divide the Cohn-Kanade database into two, 70 percent of the images are used for training and the rest for validation. The database is split several times and the accuracy is calculated for every split and the average is calculated. The ROC for the two SVM's used are shown in the figures below. The first ROC is for the SVM trained for detecting Pleasure and the second one for Arousal.

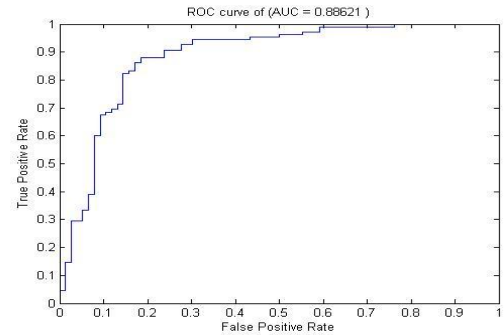


Fig. 3. ROC of the classifier for Pleasure.

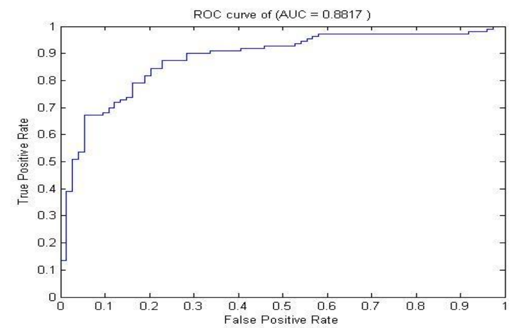


Fig. 4. ROC of the classifier for Arousal.

The accuracy of our approach over the Cohn-Kanade set is {85.32, 82.06} percent for pleasure and arousal respectively. On the other hand the approach developed by Dahmane and Meunier achieves an accuracy of only {71.80, 74.94} percent for pleasure and arousal respectively.

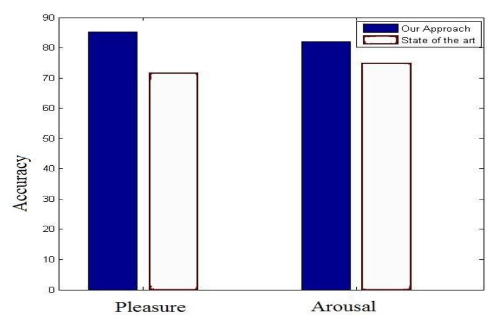


Fig. 5. Comparison of results

We also see that it takes much less time to compute Gaussian derivatives using the half-octave pyramid as compared to Gabor features because of the integer coefficient Half-Octave Pyramid used. Table 2 shows the time to calculate the features for the complete Cohn-Kanade database using the two techniques on the same machine (Intel Xeon Quad-Core 3GHz, 4GB RAM).

Table 2. Comparison of time required for calculating the two types of features

	Multi-scale Gaussian derivatives	Gabor Energy Filters
Calculation time(sec)	5.36	20.37

PCA reduces the prediction time by a factor of over 60, table 3 compares the prediction time with and without using PCA.

Table 3. Comparison of prediction time with and without PCA

	SVM with PCA	SVM without PCA
Prediction time(sec)	0.0155	0.8495

Table 4 shows the prediction time of our technique versus the state of the art because our feature vector is much smaller.

Table 4. Comparison of prediction time

	Our approach	State of the art
Prediction time(sec)	0.0155	1.06

Our approach is then tested on the FEED database and the accuracy for Pleasure-Displeasure is 70.73% while it is 70.08% for Arousal-Nonarousal.

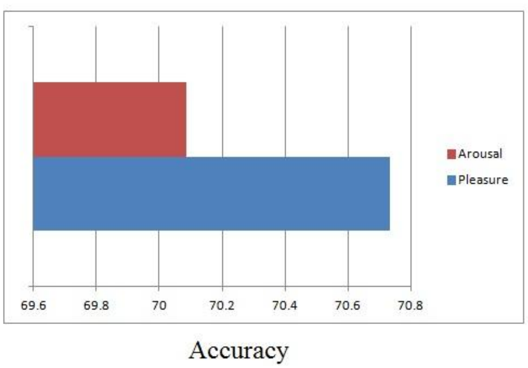


Fig. 6. Results on the FEED database

Conclusion

We have presented a novel method to analyze facial expressions and represent the underlying emotion in the PAD space. Not only is our performance better than that of the baseline approach, it is also faster at descriptor calculation and prediction. The approach performs better than the benchmark technique and is easily adaptable to mobile systems. Codes exist for calculating Multi-scale Gaussian derivatives on embedded systems using only integer coefficients.

References

1. P.Ekman. An argument for basic emotions. Cognition and Emotion, 6(3-4):169-200, 1992.
2. P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.
3. G. Littlewort, J. Whitehill, T. Wuand I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (CERT). In Proceedings of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition, pages 298–305, 2011.
4. Mohamed Dahmane and Jean Meunier. Continuous Emotion Recognition Using Gabor Energy Filters. In Affective Computing and Intelligent Interaction 2011, pages 351–358, 2011.
5. James A. Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. Journal of Research in Personality, 11(3):273–294, 1977.
6. Christian Becker-Asano. WASABI:Affect Simulation for Agents with Believable Interactivity. PhD thesis, University of Bielefeld, 2008.
7. J. Cohn T. Kanade and Y.-L. Tian. Comprehensive database for facial expression analysis. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pages 46–53, March 2000.
8. Frank Wallhoff. Facial expressions and emotion database.http://www.mmk.ei.tum.de/waf/fgnet/feed tum.html, Technical University of Munich, 2006.
9. http://www prima.inrialpes.fr/FGnet/. Face and gesture recognition working group.
10. Varun Jain and James L. Crowley. Smile detection using multi-scale gaussian derivatives. In Proceedings of the 12th WSEAS International Conference on Signal Processing, Robotics and Automation, pages 149–154, February 2013.
11. T. Lindeberg. Scale-Space Theory in Computer Vision. Kluwer Academic Press, 1994.
12. David G. Lowe. Object recognition from local scale-invariant features. In International Conference on Computer Vision, pages 1150–1157, 1999.
13. J.L. Crowley and O. Riff. Springer Lecture Notes on Computer Science, volume 2695, chapter Fast Computation of Scale normalized Gaussian Receptive Fields, pages 584–598. 2003.
14. I.T. Jolliffe. Principal Component Analysis. Springer-Verlag, 2002.
15. V.N. Vapnik. Statistical Learning Theory. Wiley, NY, 1998.
16. Paul Viola and Michael J. Jones. Robust real-time face detection. International Journal of Computer Vision, 20(17):137–154, 2004